

**Humán vizeletből dúsított O-glikopeptidek  
tandem tömegspektrometriás jellemzése  
EThcD aktiváció alkalmazásával**

Doktori (Ph.D.) értekezés tézisei

**Pap Ádám**

Témavezető:

**Dr. Darula Zsuzsanna**  
tudományos főmunkatárs

Szegedi Biológiai Kutatóközpont  
Proteomikai Kutatócsoport

Szegedi Tudományegyetem  
Természettudományi és Informatikai Kar  
Biológia Doktori Iskola

2021.

Szeged

## BEVEZETÉS

A glikoziláció a fehérjék egyik leggyakoribb poszttranszlációs módosítása. Szerepe az élő szervezetben igen sokrétű. Fehérjékhez kötődő glikánok részt vesznek különböző immunfolyamatokban, befolyásolják a fehérjék térszerkezetét, féléletidejét. Szerepük kulcsfontosságú a sejt-sejt kapcsolatokban is. A glikoziláció két fő típusa az N-, illetve O-glikoziláció. Az N-glikoziláció esetében a glikán a fehérjéhez egy Asn oldalláncának nitrogén atomján, míg az O-glikoziláció esetében egy Ser/Thr oldalláncának oxigén atomján keresztül kapcsolódik.

Az O-glikánokat a fehérjéhez közvetlenül kapcsolódó monoszacharid alapján csoportosítjuk. Emberben a leggyakoribb O-glikán típus az ún. mucin vagy GalNAc típus, amelynél a fehérjéhez közvetlenül kapcsolódó monoszacharid egy N-acetil-galaktózamin (GalNAc). A mucin típusú O-glikánok szintézise a Golgi-apparátusban történik, amely során a GalNAc-hoz további monoszacharid egységek kapcsolódnak, kialakítva a mucin típusú O-glikánok nyolc alapvázát. Minden alapváz tovább hosszabbodhat különféle struktúrákkal (pl.: N-acetil-laktózamin egységekkel, vércsoport antigénekkkel). Leggyakrabban az 1-es (Gal $\beta$ 1-3GalNAc-S/T) és 2-es (Gal $\beta$ 1-3(GlcNAc $\beta$ 1-6)GalNAc-S/T) alapvázak fordulnak elő.

A tömegspektrometria kiválóan alkalmas komplex fehérje elegyek, pl. testfolyadékok átfogó jellemzésére. Legyen szó fehérje azonosításról

vagy poszttranszlációs módosítások analiziséről, ezek jellemzése fragmentációs adatok, azaz MS/MS spektrumok alapján történik.

Az O-glikopeptidek két, kémiaiilag eltérő és fragmentáció szempontjából is különböző vegyülettípusból épülnek fel. Jellemzésükhöz egyszerre kell a peptidet és a módosító glikán(oka)t azonosítani. A poliszacharidok analízise eleve problémás, mert izomer alkotórészeik tömegük alapján nem megkülönböztethetőek. Hasonlóképpen nincs egyszerű tömegspektrometriás módszer a glikozidos kötés pontos pozíciójának és térszerkezetének jellemzésére. Tehát szerencsés esetben a következő információk nyerhetők a glikopeptidekről:

- (A) peptidszekvencia,
- (B) a glikán monoszacharid összetétele,
- (C) a glikánt alkotó monoszacharid egységek egymáshoz viszonyított helyzete,
- (D) a módosítás helye.

A tömegspektrometriás analízis során több MS/MS kísérlet szükséges különféle fragmentációs technikák alkalmazásával, hogy a két eltérő molekularészletről a lehető legtöbb szerkezeti információt szerezzük meg.

*Ioncsapda CID* esetében az egyszeres kötés-hasadások dominálnak. Glikopeptidek ioncsapda CID aktivációja a glikozidos kötések egyszeres hasadását okozza. Az ebből származó fragmensek a legintenzívebbek a spektrumban. Segítségükkel a peptidet módosító

glikánról szerkezeti információ szerezhető. Az aktivációkor keletkező deglikozilált peptid ion ( $Y_0$ ) felfedi a módosítatlan peptid tömegét, illetve a prekursor ion és az  $Y_0$  ion közti különbség a módosító glikán(ok) additív tömegét és monoszacharid összetételét is megadja. A bioszintetikus útvonalak figyelembevételével pedig a glikán(ok) monoszacharid egységeinek típusa és a glikán pontos szerkezete is kikövetkeztethető.

A *HCD aktiváció* a többszörös ütközések miatt a glikozidos kötések mellett a peptidkötés elhasadását is előidézheti, így a peptidszekvencia meghatározható a keletkező  $b$  és  $y$  szekvenciaionok segítségével. Ennél az aktivációs technikánál is megfigyelhető az  $Y_0$  ion képződése, ami itt is lehetővé teszi a peptidet módosító glikán additív tömegének és monoszacharid összetételének a meghatározását.

*Ioncsapda CID* és *HCD* adatokból nem határozható meg a módosítás helye. Mindkét aktivációs technika esetében a preferált fragmentációs esemény a gázfázisú deglikoziláció.

Az *ETD aktiváció* egy elektron-transzferrel képződött gyökion bomlásán alapuló fragmentációs technika. A fragmentáció során a peptid kötés nitrogénje és az alfa szén közötti kötés elhasadásával  $c$ , illetve  $z^•$  szekvenciaionok keletkeznek. Miután az aminosav oldalláncok épen maradnak, a megfelelő fragmensek hordozzák a módosító glikánt is, így a módosítás helye meghatározható. A glikán

szerkezetére vonatkozólag azonban az ETD spektrum nem szolgál információval.

Az *EThcD* aktiválás kombinált fragmentációs technika, amely során először az ETD, majd a teljes ionhalmazon HCD aktiváció történik. A kiegészítő HCD lépés elsősorban a már aktivált, de még el nem hasadt prekursor ionokat érinti. Az ETD lépésnek köszönhetően lehetőség nyílik a módosított peptidszekvencia azonosítására, míg a HCD aktiváció szerkezeti információval szolgál a peptidet módosító glikánról. Előnye, hogy egy spektrumból szerezhetünk információt a peptid szekvenciájáról, a módosítás helyéről és a glikán szerkezetéről.

## CÉLKITŰZÉS

A dolgozat alapjául szolgáló kutatások során humán vizeletből dúsított mucin típusú O-glikopeptidek tandem tömegspektrometriás adatainak kiértékelésére fókuszáltam. **Ehhez az alapot a jelenleg ismert legnagyobb, intakt, triptikus O-glikopeptidekről készült tömegspektrometriás adathalmaz szolgáltatta. Az adatkiértékelésnél az volt a cél, hogy jellemezzem e molekulák fragmentációs sajátosságait egy újfajta MS/MS aktivációs technika, az EThcD esetében és ezt összehasonlítsam más fragmentációs módszerekkel (ionsapda CID, HCD, ETD).**

**Az O-glikopeptidek EThcD-ben mutatott fragmentációs karakterisztikájának tanulmányozásával egyidejűleg a humán vizeletben található fehérjék O-glikozilációjának megismerése is**

**cél.** Ezek a célok nem választhatók el egymástól. A kutatás analitikai részében megszerzett ismeretek az O-glikopeptidek tandem tömegspektrometriás viselkedéséről alapjaiban meghatározzák a felvett adatokból kinyerhető, biológiailag releváns információk megbízhatóságát.

## **ALKALMAZOTT MÓDSZEREK**

### **Mintaelőkészítés**

A vizeletminták tripszinnel történő emésztése egy cellulózmembrán szűrőegységen történt. A szűrőegység alkalmazásával a mintaelőkészítéskor használt reagensek eltávolítása könnyen elvégezhető, illetve a tripszines emésztés után a keletkezett peptidek egyszerűen elválaszthatók az emésztetlen fehérjéktől. A keletkezett peptidelegyből a glikopeptidek dúsítására kétkörös búzacsíra agglutinin affinitás kromatográfiát alkalmaztam. A glikopeptidek dúsítása komplex elegyek esetében mindenképpen szükséges a tripszines emésztéskor keletkező nem glikozilált peptidek nem specifikus hátterének csökkentése érdekében.

### **LC-MS/MS analízis**

A dúsított glikopeptid-elegyeket on-line, fordított fázisú folyadékromatográfiával frakcionáltam. A tömegspektrometriás analízis során HCD adatfüggő EThcD adatgyűjtést alkalmaztam. HCD minden kiválasztott prekuzorról, míg EThcD csak a potenciális glikopeptidekről készült. Az EThcD adatgyűjtés csak akkor valósult meg, ha a kiválasztott prekuzor HCD aktivációjakor a spektrumban a

legintenzívebb 20 fragmens között volt az N-acetil-hexózamin oxónium ionja ( $m/z$  204.0867). Az EThcD mérés csak a glikopeptidek jellemzésére nélkülözhetetlen, viszont időigényesebb, mint a HCD analízis. Azaz ezzel a beállítással optimalizáltam az adatgyűjtést.

### **Adatkiértékelés**

A tömegspektrometriás adatokat többféle adatbázis kereső szoftverrel (Byonic, Protein Prospector) és manuális adatkiértékeléssel is értelmeztem. Továbbá együttműködés keretében szoftvert fejlesztettünk (GF-Hunter).

## **EREDMÉNYEK**

### **1.) EThcD aktivációval nagyobb szacharid oxónium ionok is képződhetnek**

A kétkörös WGA affinitás kromatográfiát először három egészséges donor vizeletén alkalmaztam, majd a dúsítással előállított glikopeptid-elegyek tömegspektrometriás analízisét végeztem el. Az EThcD adatokon először egy általános adatbázis keresést végeztem. A vizeletben található fehérjék O-glikozilációja kevésbé ismert, így az általános adatbázis keresés során a humán szérumban található három leggyakoribb O-glikán szerkezetet definiáltam. A keresés eredményeinek értékelése közben felfigyeltem arra, hogy az EThcD spektrumokban nagyobb szacharid oxónium ionok is képződnek az alacsony energiájú kiegészítő HCD aktiválás miatt. Ezen ionok értékes információt hordoznak arról, hogy a peptid több kisebb vagy egy nagyobb O-glikánnal módosított-e. Statisztikai analízissel

megvizsgáltam, hogy a leggyakoribb O-glikánoknak megfelelő intakt oxónium ionok milyen gyakran detektálhatók az EThcD spektrumokban. Azt találtam, hogy a nagyobb oxónium ionok, mint pl. a diszialisált core 1 szerkezetre jellemző m/z 948, vagy a core 2 hexaszacharid jelenlétére utaló m/z 1313 ionok az EThcD spektrumok zömében csak gyenge jelet adnak, így detektálásuk erősen függ a prekursor ion intenzitásától, viszont jól használhatók a glikán azonosítások megerősítésére vagy cáfolatára.

## **2.) Glikán adatbázis bővítése nem definiált módosítások segítségével**

Az általános adatbázis keresés azonosítási rátája a teljes betáplált EThcD spektrumszámhoz képest alacsony volt (~5%), amelynek két fő oka lehetséges. Az első az EThcD aktiváció alacsony hatásfoka az alacsony töltéssűrűségű prekursorok miatt. A második pedig az, hogy nagy valószínűséggel az adatbázis kereséskor nem minden, a vizeletben jelenlévő O-glikánt definiáltam. Annak érdekében, hogy feltárjam, milyen más O-glikánok módosítják a vizeletben előforduló glikopeptideket a szérumban leírt leggyakoribb szerkezeteken túl, ún. nem specifikált módosítás adatbázis keresést végeztem az EThcD spektrumokon, ahol egy bizonyos tömegtartományon belüli módosítást (0-3000 Da) engedélyeztem csakis glikozilálható aminosavakon. Az így jelentett additív tömegekből hisztogramot készítettem, és a gyakoribb tömegek valóban különböző szacharid-összetételeknek feleltek meg. Az így talált új módosítások pontosabb

jellemzését az EThcD spektrumok manuális kiértékelésével végeztem.

### **3.) Az EThcD technika segítségével meghatározható a peptidet módosító glikán monoszacharid egységeinek relatív pozíciója**

A manuális adatkiértékelés során a nem specifikált módosítás keresésből azonosított additív tömegek egy része mono- és di-O-acetil-sziálsavat, illetve disziálsavat tartalmazó O-glikánoknak felelt meg. A nagyobb szacharid oxónium ionok korábban felismert jelenléte már jelezte azt, hogy az alacsony energiájú kiegészítő HCD aktiválás glikopeptidek vizsgálatakor az egyszeres glikozidos kötéshasadásoknak kedvez. Az O-acetilált sziálsavak és disziálsavak O-glikánon belüli pozícióinak meghatározását ez a megfigyelés, azaz az egyszeres glikozidos kötéshasadások tették lehetővé. Tehát kijelenthető, hogy az alacsony energiájú kiegészítő HCD aktiválással felvett EThcD technika alkalmas a módosított peptidszekvencia mellett a peptidet módosító O-glikán szerkezeti jellemzésére. Ennek további pozitív hozadéka az, hogy **ugyanazon glikopeptid szerkezeti izomerei is megkülönböztethetők, ha az on-line fordított fázisú kromatográfia során elválnak egymástól, mint azt egy izomer pár azonosításával be is mutattam a dolgozatban.** A nem definiált módosítás kereséssel és az EThcD spektrumok manuális elemzésével **12 új O-glikán struktúrát sikerült azonosítani a vizeletben található glikoproteineken.** Ezek között, a már korábban említett szerkezeteken kívül, azonosítottam szialil-Lewis<sup>AX</sup> epitópot tartalmazó O-glikánokat is.

#### **4.) Az EThcD és HCD adatok együttes elemzése szükséges**

Az EThcD adatok mellett a HCD adatok bevonásával sikerült újabb O-glikán szerkezetekkel módosított glikopeptidek azonosítása. HCD aktiváció során a peptidet módosító glikán(ok) gázfázisú eliminációja a preferált fragmentációs esemény. Ez a HCD spektrumban a deglikozilált peptidion ( $Y_0$ ) megjelenését eredményezi. Segítségével nem csak a módosítatlan peptid molekulatömegét, de az azt módosító glikán(ok) additív tömegét is meghatározhatjuk. Az additív tömeg ismeretében pedig a monoszacharid összetétel is meghatározható. A HCD adatokból tehát előrejelezhetők EThcD adatokból már azonosított glikopeptidek új glikoformái. Ezt a hipotézist a mintákból leggyakrabban azonosított peptiddel ( $^{342}\text{AVAVTLQSH}^{350}$ , Protein YIPF3) és néhány minta adataival manuálisan teszteltem. Az eredmények meggyőzőek voltak, az adathalmaz nagysága viszont nyilvánvalóvá tette, hogy automatizálásra van szükség.

#### **5.) Szoftverfejlesztés új glikoformák azonosítására HCD adatokból**

A HCD spektrumok információtartalmának kiaknázására és a potenciális glikoformák előrejelzésének gyorsítására szoftvert fejlesztettünk (GF-Hunter), amely már a korábban megbízhatóan (pl. EThcD spektrumokból) azonosított O-glikopeptidek  $Y_0$ - $Y_1$  ionpárjait keresi a HCD spektrumokban. A minta komplexitása miatt a szűrés során egyetlen ion jelenlétére ( $Y_0$ ) támaszkodni nem elegendő. Mindenképpen szükséges az  $Y_0$  ion mellett annak  $Y_1$  (peptid+HexNAc) párját is keresni a HCD spektrumokban. Az ionpár

jelenlétének megkövetelése azonos töltésállapotban csökkenti a fals pozitív találatok arányát.

A szoftver az elfogadási paramétereknek megfelelő HCD spektrumhoz hozzárendeli a megtalált  $Y_0$ - $Y_1$  ionpárhoz tartozó peptidszekvenciát, valamint a prekursor ion és az  $Y_0$  ion tömegének felhasználásával kiszámítja a módosítás additív tömegét. Ha a módosítás tömege illik az előre megadott glikán adatbázis valamely tagjára, akkor az additív tömeget glikánként azonosítja a szoftver, melyet szükséges az adott HCD spektrum EThcD párjának manuális ellenőrzésével validálni. A szoftverbe épített glikán adatbázis iteratív módon bővíthető, így a korábban felfedezett O-glikánokat a szoftver a következő szűrés alkalmával már képes azonosítani.

#### **6.) Új glikán szerkezetek felfedezése a GF-Hunter segítségével**

A szoftverrel a tesztelést követően 10 donor vizeletéből dúsított glikopeptid-elegyek tömegspektrometriás adatait elemeztem. A HCD adatok automatizált szűrésével új oligoszacharid-kompozícióknak megfelelő additív tömegeket azonosítottam. **A szoftver segítségével további 17 új O-glikán struktúrát jellemeztem a HCD és EThcD adatok együttes felhasználásával.** Ezek között olyan új O-glikánok szerepelnek a már korábban azonosított O-acetil szíálsavat és disziálsavat tartalmazó szerkezetek mellett, amelyek **AB0 vércsoport antigént, poli-N-acetil-laktózámin egységet vagy szulfatálást tartalmaznak.**

## ÖSSZEFOGLALÁS

Létrehoztam egy hatalmas és reprezentatív tömegspektrometriás adathalmazt, amely olyan intakt humán mucin-típusú O-glikopeptidek analízisével készült, amelyeket nem laboratóriumi sejtvonalakból, hanem 10 humán donor vizeletéből izoláltam. Az adatok kiértékelése azt mutatja, hogy a vizeletben található fehérjék igen változatos glikánkészlettel módosítottak. A dolgozatban bemutatott adatkiértékelési módszerek még mindig erősen támaszkodnak a spektrumok manuális értelmezésére és az adatbázis kereső szoftverek O-glikopeptid azonosításainak felülvizsgálatára. Mindezek ellenére a vizeletben található fehérjéken **összesen 57 különböző O-glikánt** sikerült jellemezni O-glikopeptidek tandem tömegspektrometriás analízisével. **Kutatócsoportunk volt az első, amely ezeket a szerkezeteket konkrét fehérjéken, helyspecifikusan azonosította, pl. vércsoport antigének jelentését bizonyos szekretált fehérjéken. A kutatások felfedték, hogy apró különbségek az izomer glikán-szerkezetekben jelentősen befolyásolhatják a megfelelő glikoformák kromatográfiás viselkedését.**

## PUBLIKÁCIÓS LISTA

MTMT azonosító: 10058600

Összesített impakt faktor: 18.593

### **A dolgozat alapjául szolgáló közlemények:**

**Pap, A.**, Klement, E., Hunyadi-Gulyas, E., Darula, Z., & Medzihradzsky, K. (2018). Status Report on the High-Throughput Characterization of Complex Intact O-Glycopeptide Mixtures. JOURNAL OF THE AMERICAN SOCIETY FOR MASS SPECTROMETRY, 29(6), 1210–1220.  
<http://doi.org/10.1007/s13361-018-1945-7>

**IF: 3.255 (2019)**

Darula, Z., **Pap, A.**, & Medzihradzsky, K. F. (2019). Extended Sialylated O-Glycan Repertoire of Human Urinary Glycoproteins Discovered and Characterized Using Electron-Transfer/Higher-Energy Collision Dissociation. JOURNAL OF PROTEOME RESEARCH, 18(1), 280–291.  
<http://doi.org/10.1021/acs.jproteome.8b00587>

**IF: 4.074 (2019)**

**Pap, A.**, Tasnadi, E., Medzihradzsky, K. F., & Darula, Z. (2020). Novel O-linked sialoglycan structures in human urinary glycoproteins. MOLECULAR OMICS, 16(2), 156–164.  
<http://doi.org/10.1039/c9mo00160c>

**IF: 2.273 (2019)**

**További közlemények:**

**Pap, A.**, Medzihradzsky, K., & Darula, Z. (2017). Using “spectral families” to assess the reproducibility of glycopeptide enrichment: human serum O-glycosylation revisited. ANALYTICAL AND BIOANALYTICAL CHEMISTRY, 409(2), 539–550. <http://doi.org/10.1007/s00216-016-9960-7>

**IF: 3.637 (2019)**

**Pap, A.**, Prakash, A., Medzihradzsky, K. F., & Darula, Z. (2018). Assessing the reproducibility of an O-glycopeptide enrichment method with a novel software, Pinnacle. ELECTROPHORESIS, 39(24), 3142–3147. <http://doi.org/10.1002/elps.201800223>

**IF: 3.081 (2019)**

Chalkley, R. J., Medzihradzsky, K. F., Darula, Z., **Pap, A.**, & Baker, P. R. (2020). The effectiveness of filtering glycopeptide peak list files for Y ions. MOLECULAR OMICS, 16(2), 147–155. <http://doi.org/10.1039/c9mo00178f>

**IF: 2.273 (2019)**

## TÁRSSZERZŐI NYILATKOZAT I.

Alulírottak kijelentjük, hogy a közleményben (Pap, A. et al. (2018). Status Report on the High-Throughput Characterization of Complex Intact O-Glycopeptide Mixtures. JOURNAL OF THE AMERICAN SOCIETY FOR MASS SPECTROMETRY, 29(6), 1210–1220. <http://doi.org/10.1007/s13361-018-1945-7>) megjelent eredmények elérésében a jelölt, Pap Ádám, jelentős szerepet vállalt. Kijelentem továbbá azt is, hogy más jelöltnek nem adok ki hasonló tartalmú nyilatkozatot azokról az eredményekről, melyeket a jelölt fokozatszerzésében felhasznált.

.....

Dr. Klement Éva

Dr. Hunyadi-Gulyás Éva

.....

Dr. Darula Zsuzsanna

Dr. Medzihradzky F.

Katalin

Kelt: Szeged, 2021. május 28.

## TÁRSSZERZŐI NYILATKOZAT II.

Alulírottak kijelentjük, hogy a közleményben (Darula, Z., et al. (2019). Extended Sialylated O-Glycan Repertoire of Human Urinary Glycoproteins Discovered and Characterized Using Electron-Transfer/Higher-Energy Collision Dissociation. JOURNAL OF PROTEOME RESEARCH, 18(1), 280–291. <http://doi.org/10.1021/acs.jproteome.8b00587>) megjelent eredmények elérésében a jelölt, Pap Ádám, jelentős szerepet vállalt. Kijelentem továbbá azt is, hogy más jelöltnek nem adok ki hasonló tartalmú nyilatkozatot azokról az eredményekről, melyeket a jelölt fokozatszerzésében felhasznált.

.....  
Dr. Darula Zsuzsanna

.....  
Dr. Medzihradsky F.  
Katalin

Kelt: Szeged, 2021. május 28.

### **TÁRSSZERZŐI NYILATKOZAT III.**

Alulírottak kijelentjük, hogy a közleményben (Pap, A. et al. (2020). Novel O-linked sialoglycan structures in human urinary glycoproteins. MOLECULAR OMICS, 16(2), 156–164. <http://doi.org/10.1039/c9mo00160c>) megjelent eredmények elérésében a jelölt, Pap Ádám, jelentős szerepet vállalt. Kijelentem továbbá azt is, hogy más jelöltnek nem adok ki hasonló tartalmú nyilatkozatot azokról az eredményekről, melyeket a jelölt fokozatszerzésében felhasznált.

.....

Tasnádi Ervin

.....

Dr. Darula Zsuzsanna

Dr. Medzihradszky F.

Katalin

Kelt: Szeged, 2021. május 28.